

XMLTape Solution - Overview

Table of contents

1 What is the aDORe XMLTape Solution?.....	2
2 How does the aDORe XMLTape Solution work?.....	2
3 Additional Information.....	3

1. What is the aDORe XMLTape Solution?

An XMLtape is an XML file that concatenates the XML-based representation of multiple Digital Objects. The XMLtape provides a write-once/read-many XML wrapper for a collection of XML documents. The wrapper provides an easy storage format for big collections of XML files which can be processed with off the shelf tools and validated against a schema. The XMLtape is typically used in digital preservation projects.

- Integrated Indexing Interface and Implementation
- Flexible methods to access XML Documents using OAI-PMH Conventions; identifier and datestamp
- Simple Methods to extract XML Documents from XMLTape

2. How does the aDORe XMLTape Solution work?

- Write: Write multiple XML Documents to an XMLTape for ease of storage in a conventional file system.
- Index: Index key information associated with archived XML record.
- Retrieve: Extract XML Document from XMLTape.

The resulting output of the aDORe XMLTape Solution:

- XMLTape: XML-based representation of multiple Digital Objects
- XMLTape Index: Contains identifier, datestamp, length, offset, sets information about each XMLTape Record

Figure 1

In the aDORe implementation of the XMLtape, the XML-based representations of Digital Objects are DIDL documents compliant with the MPEG-21 DIDL standard. In order to keep these DIDL documents small and hence easy to process, they typically contain:

- By-Value - The metadata pertaining to the Digital Object, its constituent datastreams, and the ingestion process.
- By-Reference - The constituent datastreams of the represented Digital Object. The embedded reference in the DIDL document points to the datastream that is stored in an ARC file that is associated with the XMLtape.

The structure of XMLtapes is defined by means of an [XML Schema](#):

- An XMLtape starts off with a section that allows for the inclusion of administrative information pertaining to the XMLtape itself. Typical information includes provenance

information of the contained batch of Digital Objects, identification of the processing software, processing time, etc.

- The XMLtape-level administrative section is followed by the concatenation of records, each of which has administrative information attached to it. While allowing for the inclusion of a variety of record-level administrative information, the XMLtape has two strictly defined administrative elements: the identifier and creation datetime of the contained record. This allows for the use of a generic XMLtape processing tool that is independent of the nature of the actual included records. In aDORe, these strictly defined administrative information elements translate to the Package Identifier and the creation datetime of the DIDL document that represents a record in the XMLtape.
- The records provided in an XMLtape can be from any XML Namespace. In aDORe, they are DIDL documents compliant with the [MPEG-21 DIDL XML Schema](#).
- The XMLtape itself is a valid and well-formed XML file that can be handled by off-the-shelf XML tools for validation or parsing.

3. Additional Information

Liu, X., Balakireva, L., Hochstenbach, P., van de Sompel, H. (2005, June).

[File-based storage of Digital Objects and constituent datastreams: XMLTapes and Internet Archive ARC files](#)

Hochstenbach, P.

[XML::Tape - Perl module for the manipulation of XMLtape archives](#)